欠番オイラリアン分布とその基本統計量

城西大学　理学部　土　屋　高　宏
札幌学院大学　経済学部　中　村　永　友

概要：バケットソートを変形したソーティング・アルゴリズムに現れる離散型確率分布を導出する．このソーティングの過程で，ある規則性を持つ数が現れる．その数に関係する場合の数を漸化式で表現し，欠番がある場合の Eulerian 分布を導出する．また，漸化式を用いて基本統計量を求める．

キーワード：Eulerian 分布，モーメント，欠測，変形バケットソート

1　はじめに

1から n までの数字が書かれたよくシャッフルされている n 枚のカードが手元にある．カードを小さい順に並べ替えをするために，次の手順で行う（変形バケットソート）．(i)一番上のカードの数字が k のとき，数字 $k+1$ のカードがすでにテーブルに置かれていたらその上にカード k を載せる．(ii)もしテーブル上に $k+1$ が無ければ，どのカードの上にも載せず，テーブルにカード k を置く．(iii) (i)と(ii)を手元のカードが無くなるまで続ける．(iv)テーブルの上に m 個のカードの束ができる（$1 \leq m \leq n$）．(v)各束の一番上のカードの数字を小さい順にまとめれば，並べ替えが完了する．このとき，カードをテーブルに置き終わった時点におけるカードの束の数に関する確率分布に関心がある．土屋･中村（2009）はこの確率分布の漸化式表現を与えた．なお，「束の数」はバケットソートにおける入れ物数に対応する．

本稿では上記手順を基礎として，カードの番号に欠番がある場合の確率分布について考察する．以下では，カードの出方を（$a, b, c, \cdots$）のように表す．ただし，カードは $a, b, c, \cdots (1 \leq a, b, c, \cdots \leq n)$ の順に出現するものとし，どのカードも重複はないものとする．

2　束の数の確率分布（Eulerian 分布）

欠番のない n 枚のカードにおける束の数を表す確率変数を X とする．$X=i$ となる場合の数 $M_n(i)$ について，次の漸化式が成り立つ．

$$M_n(1)=M_n(n)=1 \quad (n\geq 1),$$
$$M_n(i)=iM_{n-1}(i)+\{n-(i-1)\}M_{n-1}(i-1) \quad (n\geq 3,\ \ i=2,\cdots,n-1).$$

この漸化式は Eulerian 数を求める式と数学的に同値であり，$M_n(i)$ と Eulerian 数 $A(n,\ i-1)$ の間に $M_n(i)=A(n,i-1)$ $(i=1,2,\cdots,n)$ が成り立つ (Euler, 1755；土屋・中村，2009)．このとき，$X=i$ である確率 $P_n(i)$は，$P_n(i)=M_n(i)/n!$ で定義され，次の漸化式で与えられる．

$$P_n(1)=P_n(n)=\frac{1}{n!} \quad (n\geq 1),$$
$$P_n(i)=\frac{i}{n}P_{n-1}(i)+\frac{n-(i-1)}{n}P_{n-1}(i-1) \quad (n\geq 3,\ \ i=2,\cdots,n-1). \tag{1}$$

この漸化式で与えられる確率分布を Eulerian 分布（土屋・中村，2009）と呼ぶ（図 1 参照）．

3　モーメントと漸化式

Eulerian 分布 (1) の r 次モーメント $E_n(X^r)$ は，以下の漸化式で与えられる（土屋・中村，2009)．

$$E_1(X^r)=1,\ \ E_n(X^r)=\sum_{k=0}^{r}\left\{\frac{n+1}{n}\binom{r}{k}-\frac{1}{n}\binom{r+1}{k}\right\}E_{n-1}(X^k) \quad (n\geq 2). \tag{2}$$

(2) 式より，確率分布 (1) の期待値と分散は，それぞれ次式で与えられる．

$$E_n(X)=\frac{n+1}{2} \quad (n\geq 1), \qquad V_1(X)=0,\ \ V_n(X)=\frac{n+1}{12} \quad (n\geq 2).$$

4　欠番がある場合の Eulerian 分布

前述のソーティング過程を一般に拡張し，カードの番号に欠番がある場合の確率分布について考察する．

いま，1 から n までの数字が書かれた n 枚のカードのうち1枚だけ欠番がある状況を考える．どの番号も等確率で欠番が生じるという仮定の下で，束の数 X の確率分布を導出する．ある特定の番号が欠番している場合も興味の対象であるが，本稿では取り扱わない．

例えば $n=3$ の場合，欠番がなければカードの出方は (1，2，3)，(1，3，2)，(2，1，3)，(2，3，1)，(3，1，2)，(3，2，1) の 6 通りである．1 枚だけ欠番があるという状況は 3 番目のカードが欠番していることと同値なので，次の 6 通りとなる：(1，2)，(1，3)，(2，1)，(2，3)，(3，1)，(3，2)．そして，番号 3 が欠番のとき，すなわち (1，2)，(2，1) のときだけ，束が 1 つ減少する．それ以外は束の数に変化はない．$n=4$ のときも同様であり，番号 4 が欠番のときだけ，束が 1 つ減少する．この配列は $n=3$ の配列に等しいことから，欠番がある場合の束の数の総数は次のように得られる．$n=3$ の場合は番号 3 が欠番のときだけ，束が 1 つ減少し，束の数 3 が 2 に，2 が 1 になるので，欠番がある場合に束の数が i となる総数 $M_3^*(i)$ は欠番なしの束の総数 $M_3(i)$ に $n=2$ のときの総数 $M_2(i)$ をたして，それと同数だけ減少した数 $M_2(i-1)$ を引くことにより，求めることができる．$n=4$ の場合は，$M_4^*(i)=M_4(i)+M_3(i)-M_3(i-1)$ となる．

以上の考え方に基づき，n 枚のカードにおける束の数 $X=i$ となる総数 $M_n^*(i)$ について，次の漸化式が成り立つ．

$$M_n^*(1)=2 \quad (n\geq 2),\quad M_n^*(n)=0 \quad (n\geq 1),$$
$$M_n^*(i)=M_n(i)+M_{n-1}(i)-M_{n-1}(i-1) \quad (n\geq 3,\quad i=2,\cdots,n-1).$$

ただし，$M_n(i)$ は欠番なしの束の数を表す．このとき，$X=i$ である確率 $P_n^*(i)$ は，$P_n^*(i)=M_n^*(i)/n!$ で定義され，次の漸化式で与えられる．

$$P_n^*(1)=\frac{2}{n!} \ (n\geq 2),\quad P_n^*(n)=0 \quad (n\geq 1),$$
$$P_n^*(i)=\frac{i+1}{n}P_{n-1}(i)+\frac{n-i}{n}P_{n-1}(i-1) \quad (n\geq 3,\quad i=2,\cdots,n-1). \tag{3}$$

ただし，$P_n(i)$ は(1)式を満たすものとする．

図 2 に束の数 $M_n^*(i)$ の確率分布 $P_n^*(i)$ の例を示す．

定理 4.1　確率分布(3)の r 次モーメント $E_n^*(X^r)$ は，以下の漸化式で与えられる．

$$E_1^*(X^r)=0,\quad E_n^*(X^r)=E_n(X^r)-\frac{1}{n}\sum_{k=0}^{r-1}\binom{r}{k}E_{n-1}(X^k) \quad (n\geq 2). \tag{4}$$

ただし，$E_n(X^r)$ は(2)式を満たすものとする．確率分布(3)の期待値と分散は，それぞれ次式で与えられる．

$$E_n^*(X)=\frac{n+1}{2}-\frac{1}{n} \quad (n\geq 1),\qquad V_1^*(X)=0,\quad V_n^*(X)=\frac{n+1}{12}-\frac{1}{n^2} \quad (n\geq 2). \tag{5}$$

証明　$n=1$ のとき，$E_1^*(X^r)=1^rP_1^*(1)=0$ である．$n\geq 3$ のとき，(3)式から

$$\sum_{i=2}^{n-1}i^rP_n^*(i)=\frac{1}{n}\sum_{i=2}^{n-1}(i^{r+1}+i^r)P_{n-1}(i)+\frac{1}{n}\sum_{i=2}^{n-1}(ni^r-i^{r+1})P_{n-1}(i-1)$$

となる．$i=1$ のとき，$1^rP_n^*(1)=2/n!$ であり，$i=n$ のとき，$n^rP_n^*(n)=0$ であるから，

$$\sum_{i=1}^{n}i^rP_n^*(i)=\frac{1}{n}\sum_{i=1}^{n-1}(i^{r+1}+i^r)P_{n-1}(i)+\frac{1}{n}\sum_{i=1}^{n-1}(n-i-1)(i+1)^rP_{n-1}(i),$$

すなわち，次式が成り立つ．

$$E_n^*(X^r)=\frac{1}{n}E_{n-1}(X^{r+1})+\frac{1}{n}E_{n-1}(X^r)+\frac{1}{n}E_{n-1}[(n-X-1)(X+1)^r]. \tag{6}$$

(6)式の右辺第 3 項は

$$\begin{aligned}&\frac{1}{n}E_{n-1}[(n-X-1)(X+1)^r]\\&=E_{n-1}[(X+1)^r]-\frac{1}{n}E_{n-1}[X(X+1)^r]-\frac{1}{n}E_{n-1}[(X+1)^r]\\&=\sum_{k=0}^{r}\binom{r}{k}E_{n-1}(X^k)-\frac{1}{n}\sum_{k=0}^{r}\binom{r+1}{k}E_{n-1}(X^k)-\frac{1}{n}E_{n-1}(X^{r+1})\end{aligned}$$

と変形できるので，(6)式は

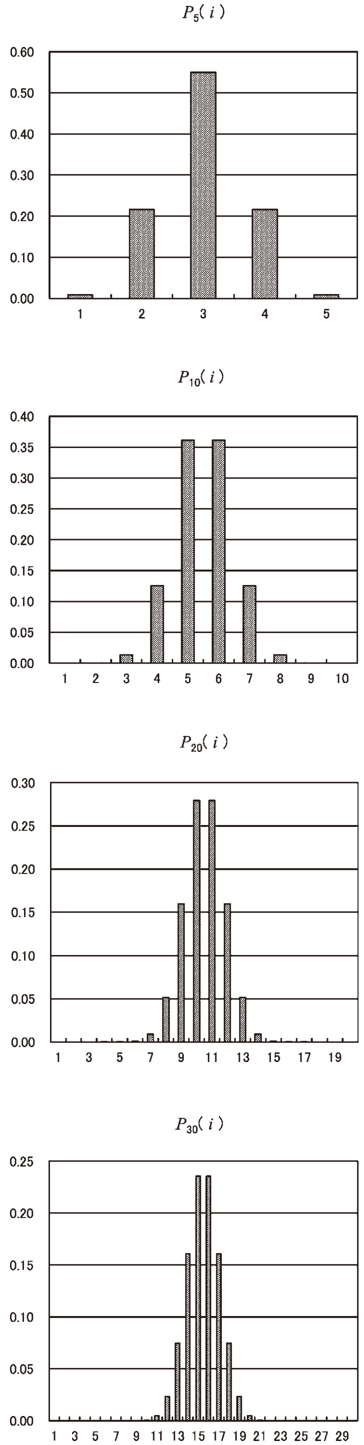

図 1：$P_n(i)(n=5, 10, 20, 30)$ の確率分布

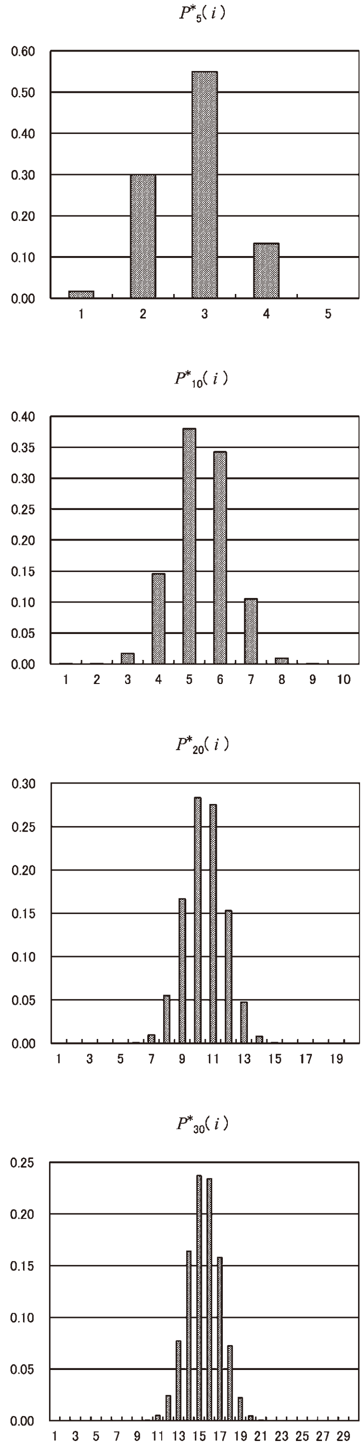

図 2：$P^*_n(i)(n=5, 10, 20, 30)$ の確率分布

$$E_n^*(X^r)=\sum_{k=0}^{r}\left\{\binom{r}{k}-\frac{1}{n}\binom{r+1}{k}\right\}E_{n-1}(X^k)+\frac{1}{n}E_{n-1}(X^r) \tag{7}$$

となる．$n=2$ のとき，(3)式から $E_2^*(X^r)=1^rP_2^*(1)+2^rP_2^*(2)=1$ となるから，(7)式は $n=2$ のときも成り立つ．ここで，(2)式を(7)式に代入することにより，

$$E_n^*(X^r)=E_n(X^r)-\frac{1}{n}\sum_{k=0}^{r-1}\binom{r}{k}E_{n-1}(X^k)$$

を得る．

$r=1$ のとき，

$$\sum_{k=0}^{r-1}\binom{r}{k}E_{n-1}(X^k)=E_{n-1}(1)=1$$

より，期待値は

$$E_n^*(X)=E_n(X)-\frac{1}{n}=\frac{n+1}{2}-\frac{1}{n}$$

となる．$r=2$ のとき，

$$\sum_{k=0}^{r-1}\binom{r}{k}E_{n-1}(X^k)=1+2E_{n-1}(X)=1+n$$

より，

$$E_n^*(X^2)=E_n(X^2)-\frac{1}{n}(1+n)$$

となるから，分散は

$$\begin{aligned} V_n^*(X)&=E_n(X^2)-\frac{1}{n}(1+n)-\left\{E_n(X)-\frac{1}{n}\right\}^2\\ &=V_n(X)-\frac{1}{n^2}\\ &=\frac{n+1}{12}-\frac{1}{n^2} \end{aligned}$$

となる． □

5　数値実験

表1は1から n までの数字からいずれか1つを等確率に欠番させて，標本平均値と不偏分散を求めた結果と(5)式を用いて求めた期待値，分散（理論値）を比較したものである．数値実験の結果は理論値とほぼ一致していることがわかる．

表１．理論値との比較

	数値実験		理論値	
n	標本平均	不偏分散	期待値	分散
5	2.80	0.450	2.80	0.460
10	5.40	0.911	5.40	0.907
20	10.45	1.762	10.45	1.748
50	25.48	4.249	25.48	4.250
100	50.53	8.486	50.49	8.417

6　おわりに

本稿ではカード番号に欠番がある場合の Eulerian 分布とそのモーメントを導出した．確率分布は非対称となり，期待値と分散については，欠番がない場合とのズレがそれぞれ $1/n$ と $1/n^2$ であることがわかった．今後の課題として，他の確率分布との関係や近似の精密化，複数の欠番がある場合に確率分布を導出することなどが挙げられる．

参考文献

Euler, L. (1755). *Institutiones Calculi Differentialis*, Petrograd.

土屋高宏，中村永友（2007）．ある種の並べ替え算法における分布について，札幌学院大学商経論集，**24**(2)，31-47．

土屋高宏，中村永友（2009）．変形バケットソートに現れる離散型確率分布と Eulerian 数，統計数理，**57**(1)，31-50．

Eulerian Distribution with a Missing Number

by

Takahiro TSUCHIYA[1] and Nagatomo NAKAMURA[2]

Abstract

A discrete distribution induced by the sorting algorithm of modified bucket sort is proposed. The systematic numbers appear in this sorting process. The recurrence relation for Eulerian distribution with a missing number is given. The moments of the distribution is also obtained.

Keywords: Eulerian Distribution, Moments, Missing Number, Modified Bucket Sorting

[1]Department of Mathematics, Josai University, takahiro@math.josai.ac.jp
[2]Department of Economics, Sapporo Gakuin University, nagatomo@sgu.ac.jp